Title

Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study.

Journal

Nature genetics, 41(2)

ISSN

1061-4036

Authors

Silverberg, Mark S
Cho, Judy H
Rioux, John D
et al.

Publication Date

2009-02-01

DOI

10.1038/ng.275

Peer reviewed

HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2009 August 01.

Published in final edited form as:

Nat Genet. 2009 February ; 41(2): 216–220. doi:10.1038/ng.275.

Ulcerative colitis loci on chromosomes 1p36 and 12q15 identified by genome-wide association study

Mark S. Silverberg[*],
msilverberg@mtsinai.on.ca, Mount Sinai Hospital IBD Group, University of Toronto, 600 University Avenue, Toronto, ON M5G1X5, Canada.

Judy H. Cho[*],
judy.cho@yale.edu, Section of Digestive Diseases, Departments of Medicine and Genetics, Yale University, 333 Cedar Street, LMP1080, New Haven, CT 06520, USA.

John D. Rioux,
john.david.rioux@umontreal.ca, Université de Montréal and the Montreal Heart Institute, Research Center, 5000 rue Belanger, Montreal, QC H1T1C8, Canada.

Dermot P.B. McGovern,
Dermot.McGovern@cshs.org, Inflammatory Bowel Disease Center, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA.

Jing Wu,
jwu@stat.cmu.edu, Department of Statistics, Carnegie Mellon University, Baker/Porter Hall A60K, Pittsburgh, PA 15213, USA.

Vito Annese,
v.annese@operapadrepio.it, IRCCS -CSS Hospital, Viale Cappuccini, 1, S. Giovanni Rotondo, Fg 71013, Italy.

Jean-Paul Achkar,
achkarj@ccf.org, Digestive Disease Institute, Cleveland Clinic, 9500 Euclid Avenue, Desk A31, Cleveland, OH 44195, USA.

Philippe Goyette,
philippe.goyette@inflammgen.org, Université de Montréal and the Montreal Heart Institute Research Center, 5000 rue Bélanger, Montreal, QC H1T1C8, Canada.

[#]corresponding author.
[*]indicates equal contributions

Author contributions

J.H.C., E.O.K., and L.P.S. developed and maintained the NIDDK IBD Genetics Consortium Data Coordinating Center infrastructure. C.A., J-P.A., V.A., T.M.B., F.B., S.R.B., J.H.C., R.H.D., A.M.G., A.F.I., R.G.L., A.L., D.P.B.M., P.P., D.D.P., M.D.R., J.D.R., J.I.R., R.S., M.S.S., A.H.S., S.R.T., and K.D.T. provided patient samples and clinical information. S.R.B., J.H.C., M.J.D., R.H.D., P.K.G., A.T.L., J.D.R., J.I.R., M.S.S., and K.D.T. designed the GWAS. M.M.B. and R.H.D. performed quality control and preliminary association analyses of the GWAS data. J.W. performed the GEM, quantile-quantile, and conditional analyses of the GWAS data under the supervision of K.R., with contributions from L.K. M.J.D. defined the “best region” SNPs among SNPs with GWAS P < 0.0001. W.X. identified the best GWAS proxies for CD and UC loci. R.H.D., P.G., J.D.R., and R.S. designed and performed the replication study. R.H.D. and P.G. analyzed the replication data. The manuscript was written by J.H.C., R.H.D., K.R. and M.S.S. with contributions from C.A., M.M.B., S.R.B., P.K.G., D.P.B.M., J.D.R., J.I.R., R.S., and J.W. R.H.D. coordinated the genotyping, analysis and manuscript writing efforts of this multicenter study.

Regan Scott,
res45@pitt.edu, Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, UPMC-PUH, Mezzanine Level, C-Wing, 200 Lothrop Street, Pittsburgh, PA 15213, USA.

Wei Xu,
wxu@uhnres.utoronto.ca, Princess Margaret Hospital, Department of Public Health Sciences, University of Toronto, 610 University Avenue, Toronto, ON M5G2M9, Canada.

M. Michael Barmada,
barmada@pitt.edu, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, 130 Desoto Street, Pittsburgh, PA 15261, USA.

Lambertus Klei,
kleil@upmc.edu, University of Pittsburgh Medical Center, Department of Psychiatry, 3811 O'Hara Street, Pittsburgh, PA 15213-2593, USA.

Mark J. Daly,
mjdaly@chgr.mgh.harvard.edu, Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, MA 02114, USA.

Clara Abraham,
clara.abraham@yale.edu, Section of Digestive Diseases, Department of Medicine, Yale University, 333 Cedar Street, LMP1080, New Haven, CT 06520, USA.

Theodore M. Bayless,
tbayless@jhmi.edu, Johns Hopkins University School of Medicine, Department of Medicine, Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center, 600 N. Wolfe Street, Baltimore, MD 21287, USA.

Fabrizio Bossa,
fabrizio.bossa@libero.it, IRCCS - CSS Hospital, Viale Cappuccini, 1, S. Giovanni Rotondo, Fg 71013, Italy.

Anne M. Griffiths,
anne.griffiths@sickkids.ca, The Hospital for Sick Children, Department of Pediatrics, 555 University Avenue, Toronto, ON M5G1X8, Canada.

Andrew F. Ippoliti,
Andrew.Ippoliti@cshs.org, Inflammatory Bowel Disease Center, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA.

Raymond G. Lahaie,
lahaier@sympatico.ca, Department of Medicine, Université de Montréal and Hopital Saint-Luc, 1058, rue Saint-Denis, Montréal, QC H2X3J4, Canada.

Anna Latiano,
a.latiano@operapadrepio.it, IRCCS - CSS Hospital, Viale Cappuccini, 1, S. Giovanni Rotondo, Fg 71013, Italy.

Pierre Paré,

pierre.pare.cha@ssss.gouv.qc.ca, Division of Gastroenterology, Department of Medicine, Laval University, Quebec City, CHAUQ – Hôpital St. Sacrement, 1050, Chemin Ste. Foy, QC G1S4L8, Canada.

Deborah D. Proctor,
deborah.proctor@yale.edu, Yale University, Department of Medicine, 333 Cedar Street, LMP 1080, New Haven, CT 06520, USA.

Miguel D. Regueiro,
mdr7@pitt.edu, Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, UPMC-PUH, Mezzanine Level, C-Wing, 200 Lothrop Street, Pittsburgh, PA 15213, USA.

A. Hillary Steinhart,
hsteinhart@mtsinai.on.ca, Mount Sinai Hospital IBD Group, University of Toronto, 600 University Avenue, Toronto, ON M5G1X5, Canada.

Stephan R. Targan,
Stephan.Targan@cshs.org, Inflammatory Bowel Disease Center, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA.

L. Philip Schumm,
pschumm@uchicago.edu, Department of Health Studies, University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637, USA.

Emily O. Kistner,
ekistner@uchicago.edu, Department of Health Studies, University of Chicago, 5841 S. Maryland Avenue, Chicago, IL 60637, USA.

Annette T. Lee,
anlee@nshs.edu, The Feinstein Institute for Medical Research, 350 Community Drive, Manhasset, NY 11030, USA.

Peter K. Gregersen,
peterg@nshs.edu, The Feinstein Institute for Medical Research, 350 Community Drive, Manhasset, NY 11030, USA.

Jerome I. Rotter,
Jerome.Rotter@cshs.org, Medical Genetics Institute and Inflammatory Bowel Disease Center, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA.

Steven R. Brant,
sbrant@jhmi.edu, Johns Hopkins University School of Medicine, Department of Medicine, Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center; and Johns Hopkins University Bloomberg School of Public Health, Department of Epidemiology; 1501 E. Jefferson Street, B136, Baltimore, MD 21231, USA.

Kent D. Taylor,
Kent.Taylor@cshs.org, Medical Genetics Institute and Inflammatory Bowel Disease Center, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA.

Kathryn Roeder, and

roeder@stat.cmu.edu, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA.

Richard H. Duerr[#]

duerr@pitt.edu, Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine; and Department of Human Genetics, University of Pittsburgh Graduate School of Public Health; UPMC-PUH, Mezzanine Level, C-Wing, 200 Lothrop Street, Pittsburgh, PA 15213, USA.

Abstract

Ulcerative colitis is a chronic inflammatory disease of the colon that presents as diarrhea and gastrointestinal bleeding. We performed a genome-wide association study using DNA samples from 1,052 individuals with ulcerative colitis and pre-existing data from 2,571 controls, all of European ancestry. In an analysis that controlled for gender and population structure, ulcerative colitis loci attaining genome-wide significance and subsequent replication in two independent populations were identified on chromosomes 1p36 (rs6426833, combined $P = 5.1\times10^{-13}$, combined OR = 0.73) and 12q15 (rs1558744, combined $P = 2.5\times10^{-12}$, combined OR = 1.35). In addition, combined genome-wide significant evidence for association was found in a region spanning *BTNL2* to *HLA-DQB1* on chromosome 6p21 (rs2395185, combined $P = 1.0\times10^{-16}$, combined OR = 0.66) and at the *IL23R* locus on chromosome 1p31 (rs11209026, combined $P = 1.3\times10^{-8}$, combined OR = 0.56; rs10889677, combined $P = 1.3\times10^{-8}$, combined OR = 1.29).

Ulcerative colitis (UC) and Crohn's disease (CD) are the two major forms of inflammatory bowel disease (IBD). Extensive evidence suggests that IBD occurs when ubiquitous commensal enteric bacteria initiate and perpetuate a dysregulated mucosal immune response in genetically susceptible individuals1. CD and UC share many characteristics but unique clinical features also distinguish them, suggesting that they share some genetic susceptibility loci but differ at others.

Genome-wide association studies (GWAS) in CD and a CD GWAS meta-analysis provide genome-wide significant evidence for 32 CD loci2–7. The most significant CD-associated loci include *IL23R, NOD2, ATG16L1* and a gene desert on chromosome 5p13, implicating both the innate and acquired immune responses. A non-synonymous SNP scan found association between UC and *ECM1* on chromosome 1q218. The same study and another also reported association between UC and some CD loci8,9. A GWAS in a pediatric-onset IBD sample identified general IBD loci on chromosomes 20q13 and 21q2210. A UC GWAS in European samples found combined, genome-wide significant evidence for association with *IL10* and other suggestive associations11. There are many reports of UC association with HLA Class II alleles12 and major histocompatibility complex (MHC) SNPs8,10,11.

We conducted a GWAS that included 1,052 individuals with UC extending beyond the rectum and 2,571 controls, all of white, non-Hispanic, European ancestry. Principal components of ancestry analysis demonstrated population structure (see Supplementary Fig. 1 online). We used genetic matching (GEM) within gender to control for structure and potential effects of gender, followed by allelic association tests by conditional logistic regression on gender-ancestry strata13 (see Supplementary Table 1 online). GEM reduces

the false positive rate attributable to ancestry mismatches between cases and controls13. The genomic control inflation factor14 in the combined GEM results was 1.04 compared to 1.17 for pre-GEM traditional chi square tests for allelic association. Quantile-quantile plots showed deviation from expectation at a substantially higher chi-square value in the GEM dataset (see Supplementary Fig. 2b online) compared to the pre-GEM dataset (see Supplementary Fig. 2a online). A second stage of our study aimed to replicate association in two regions that showed genome-wide significance and to extend the association evidence at other independent loci with GWAS P values less than 1×10^{-4} (see Supplementary Table 2 online). P less than 1×10^{-4} was the chosen threshold for our follow-up studies because the corresponding chi-square value (15.136) is just above the point where the quantile-quantile plot of GEM results began to deviate from expectation (see Fig. 2b online). Two independent replication samples included 768 UC cases and 721 controls from North America, all of white, non-Jewish, non-Hispanic, European ancestry, and 619 UC cases and 394 controls from southern Italy after quality control (see Supplementary Table 1 online).

Genome-wide significant association in the GWAS with subsequent replication was observed on chromosome 1p36 (see Table 1 and Figure 1a) at the correlated ($r^2 = 0.76$) markers rs6426833 (GWAS P = 6.8×10^{-10}, GWAS and replication combined P = 5.1×10^{-13}) and rs10753575 (GWAS P = 2.0×10^{-9}, GWAS and replication combined P = 9.4×10^{-11}). A residual association signal was observed for rs6426833 but not rs10753575 when these SNPs were paired in a conditional analysis of their GWAS data (see Supplementary Table 3 online). A third SNP in this region, rs3806308, also showed genome-wide significant association in the GWAS with subsequent replication (GWAS P = 4.7×10^{-8}, GWAS and replication combined P = 6.7×10^{-9}); rs3806308 was not correlated with either rs6426833 or rs10753575 ($r^2 < 0.01$) and residual association signals were observed when rs3806308 and each of the other two SNPs were paired in conditional analyses of their GWAS data (see Supplementary Table 3 online). Therefore, the chromosome 1p36 locus has at least two independent association signals (rs6426833 and rs3806308) separated by recombination hotspots (see Figure 1a). There was also significant transmission/disequilibrium test (TDT)15 evidence for under-transmission of the rs6426833 G allele (transmitted:untransmitted = 197:239, P = 4.4×10^{-2}) in 436 trios of UC GWAS or replication cases and their parents.

On chromosome 12q15 (see Table 1 and Figure 1d), genome-wide significant evidence for association and subsequent replication was observed at rs1558744 (GWAS P = 5.5×10^{-10}, GWAS and replication combined P = 2.5×10^{-12}). rs7134599, which was correlated with rs1558744 ($r^2 = 0.92$), showed association evidence equivalent to the evidence for rs1558744 in a conditional analysis of the GWAS data (see Supplementary Table 3 online) and genome-wide significant combined evidence for association in the GWAS and replication samples (GWAS P = 5.4×10^{-7}, GWAS and replication combined P = 6.0×10^{-9}). rs2870946, which was only weakly correlated with rs1558744 and rs7134599 ($r^2 = 0.031$ and 0.029, respectively), showed suggestive combined evidence for association in the GWAS and replication samples that was close to the threshold for genome-wide significance (GWAS P = 1.0×10^{-5}, GWAS and replication combined P = 4.8×10^{-7}), and residual association signals were observed when rs2870946 and each of the other two SNPs were paired in conditional analyses of the GWAS data (see Supplementary Table 3 online).

Therefore, the chromosome 12q15 locus appears to have two independent association signals (rs1558744 and rs2870946) separated by recombination hotspots (see Figure 1d). The rs1558744 A allele also showed TDT15 evidence for association with UC (transmitted:untransmitted = 238:180, $P = 4.6\times10^{-3}$).

The GWAS and replication samples showed combined genome-wide significant evidence for association at two additional loci which have been implicated in UC previously, namely *IL23R* (interleukin 23 receptor) on chromosome 1p312 (see Table 1 and Figure 1b) and the MHC on chromosome 6p218,10–12 (see Supplementary Table 2 online, Table 1, and Figure 1c). Conditional analyses of the UC GWAS data suggest that there are at least two independent association signals in the *IL23R* region, consistent with our earlier study which identified the CD- and UC-associated *IL23R* locus2, and multiple independent association signals in the MHC (see Supplementary Table 3). The most significantly associated SNPs at the *IL23R* and MHC loci also showed TDT15 evidence for association with UC (rs111209026 A allele, transmitted:untransmitted = 25:63, $P = 5.1\times10^{-5}$; rs10889677 A allele, transmitted:untransmitted = 223:150, $P = 1.6\times10^{-4}$; rs2395185 A allele, transmitted:untransmitted = 121:224, $P = 2.9\times10^{-8}$). See Supplementary Note online for further discussion about the association evidence at rs2395185.

OTUD3 (OTU domain containing 3) and part of *PLA2G2E* (phospholipase A2, group IIE) are located within an approximately 100 kb region containing the most significant chromosome 1p36 association signal (rs6426833, see Figure 1a). *OTUD3* is expressed broadly and has homology to an OTU-like cysteine protease16. *PLA2G2E* is a compelling candidate, as the sPLA2 family of proteins releases arachidonic acid from membrane phospholipids which leads to the production of proinflammatory lipid mediators, such as prostaglandins and leukotrienes17. Furthermore, *PLA2G2E* expression in the lung and small intestine is induced with lipopolysaccharide stimulation, suggesting a role in bacterial-associated inflammation17,18. The second independent association signal on chromosome 1p36 (rs3806308, see Figure 1a) is located within 1.1 kb of *RNF186* (ring finger protein 186). Ring finger proteins are involved in ubiquitination of proteins and diverse cellular processes19.

The most significant chromosome 12q15 association signal (rs1558744, see Figure 1d) is located in a region that is devoid of established coding genes, but the *IFNG* (interferon, gamma; IFNγ), *IL26* (interleukin 26), and *IL22* (interleukin 22) genes are located 44 kb, 91 kb, and 137 kb, respectively, telomeric to rs1558744. A second, highly suggestive association signal at rs2870946 is independent of rs1558744 and located in *IL26* (see Figure 1d). IFNγ is critical in the immune response to pathogens, in part through regulation of macrophage function; it regulates numerous levels of immune homeostasis including T cell subsets, NK cells and NK T cells20. *IL22* and *IL26* are secreted by Th17 cells, which mediate host defense to infections as well as tissue inflammation in many chronic, immune-mediated diseases, including IBD21. In addition to a role in mediating host defense to bacterial pathogens21, IL-22 can mediate protection during acute inflammation, including intestinal inflammation22.

The chromosome 6p21 association signal maps to a broad region (see Supplementary Table 2 online) that spans many genes in the MHC and demonstrates extensive linkage disequilibrium. However, the maximal association signal at rs2395185 is located in an approximately 300 kb interval spanning the *BTNL2* (butyrophilin-like 2), *HLA-DRA*, *HLA-DRB5*, *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1* genes (see Figure 1c). This region was also implicated in other studies8,10–12. rs9268877 showed the most significant evidence for association with UC in a study by Franke et al.11 and is correlated ($r^2 = 0.61$) in HapMap CEU data23 with rs6903608, which was the second most significantly associated MHC SNP in our UC GWAS (see Supplementary Table 2 online). rs2516049, which was correlated with rs2395185 ($r^2 = 0.80$) and showed association evidence equivalent to that for rs2395185 in our UC GWAS data (see Supplementary Table 3 online), demonstrated association that was nearly as significant as the evidence for the maximally associated MHC SNP in a study by Kugathasan et al10. The rs2395185 risk allele was associated with increased average expression of *HLA-DRB1* and *HLA-DQA1* in a lymphoblastoid cell line eQTL analysis24.

Of the remaining loci tested in our replication samples (see Supplementary Table 2 online), a cluster of SNPs on chromosome 7q31 flanked on one side by *SLC26A3* (solute carrier family 26, member 3) or located within *DLD* (dihydrolipoamide dehydrogenase) or *LAMB1* (laminin, beta 1) on the other side showed promising extension of the evidence for association (rs4730276, combined $P = 9.0\times10^{-6}$; rs4598195, combined $P = 9.6\times10^{-7}$; rs2158836, combined $P = 6.7\times10^{-6}$). *LAMB1* is a plausible functional candidate since laminins play a role in intestinal health and disease25.

Finally, we determined whether CD loci implicated in a recent meta-analysis and CD or UC loci reported in other studies showed evidence for association in our UC GWAS (see Supplementary Table 4 online). When we assessed the best UC GWAS proxy for each previously implicated SNP, we observed only nominally ($P < 0.05$) significant (see Supplementary Table 4a online) or no significant (see Supplementary Table 4b online) evidence for association with these loci in our UC GWAS and no significant evidence for association for some of these loci that we tested in the replication samples (see Supplementary Tables 4a and 4b online).

In this study, we identified UC loci on chromosomes 1p36 and 12q15 where compelling candidate genes include *PLA2G2E*, *IFNG*, *IL26* and *IL22*. We also confirmed association between UC and the MHC region, where associated SNPs are correlated with altered HLA Class II expression, and between UC and *IL23R*. This report extends the model suggesting that there exist both unique as well as shared genetic factors for UC and CD. It is highly likely that additional UC loci remain to be found since our study had limited power to detect small effect sizes, especially for lower risk allele frequencies (see Supplementary Table 5 online)26. The experience in CD and other complex traits is that individual studies with sample sizes similar to the present study sample are sufficiently powered to detect risk loci with larger effects but find only a fraction of loci with smaller effects; loci with smaller effects can be detected via meta-analysis of the individual studies7. The loci identified through GWAS may eventually establish important new therapeutic targets for the treatment of IBD.

METHODS

UC GWAS study subjects, genotyping and analysis

1,052 white, non-Hispanic, European ancestry patients with UC extending beyond the rectum and 2,571 white, European ancestry controls were genotyped using Illumina HumanHap300v1, HumanHap300v2, HumanHap550v1 or HumanHap550v3 Genotyping BeadChips (Illumina, Inc., San Diego, CA,) at the Feinstein Institute for Medical Research. Control data were obtained from our CD GWAS2,4 or from studies 64 and 65 deposited in the Illumina iControlDB (http://www.illumina.com/pages.ilmn?ID=231). Informed consent was obtained using protocols approved by each local institutional review board. More details about study subjects and the Illumina Genotyping BeadChips used to genotype groups of study subjects are available in the Supplementary Methods and Supplementary Table 1 online.

Data management and quality control filtering are described in the Supplementary Methods online. Case-control matching, by gender and ancestry, followed by association analyses by conditional logistic regression on gender-ancestry strata in the matched datasets (see Supplementary Table 1 online) were performed using GEM13. The GEM analysis is described in more detail in the Supplementary Methods online. The genomic control inflation factor was computed14. SNPs with nominal P values $< 5\times10^{-8}$ were considered to have genome-wide significant evidence for association27.

Replication study subjects, genotyping and analysis

Two independent replication samples were included in the second stage of our study (see Supplementary Table 1 online). A North American replication sample consisted of 769 patients with UC and 727 controls, all of white, non-Jewish, non-Hispanic, European ancestry. The second replication sample consisted of 633 patients with UC and 415 controls from southern Italy.

SNPs with UC GWAS P values $< 1\times10^{-4}$ were included on a list of replication genotyping candidates with priority scores for inclusion in the replication genotyping assigned according to significance of the association evidence. Among correlated ($r^2 > 0.5$) SNPs with UC GWAS P values $< 1\times10^{-4}$, only the SNP with the most significant P value was chosen as the “best region” SNP and placed on the replication candidate list. Other replication genotyping candidates included a) SNPs at previously implicated CD or UC loci or their best UC GWAS proxies ($r^2 > 0.5$ in a 1 Mb window of HapMap CEU data23 centered on the implicated SNP) that showed nominal evidence for association ($P < 0.05$) in our UC GWAS, and b) previously implicated SNPs that did not have a proxy in our UC GWAS. Finally, additional SNPs from loci that demonstrated genome-wide significant evidence for association were added to the replication candidate list. rs6426833, rs2395185 and rs1558744 were genotyped by Pyrosequencing (Biotage AB, Uppsala, Sweden) at the University of Pittsburgh. SNPs on the replication candidate list that had designable Sequenom iPlex (Sequenom, San Diego, CA) assays and could be multiplexed in three oligonucleotide pools were genotyped on the Sequenom MassArray platform at the University of Pittsburgh (North American samples and some of the Italian sample

genotyping) and the Montreal Heart Institute (Italian samples). Samples with high genotyping failure rates were attempted a second time. Data management and quality control filtering are described in the Supplementary Methods online. Association evidence in the two replication samples was assessed using the Cochran-Mantel-Haenszel (CMH) test, and the Fisher method was used to combine UC GWAS GEM P values and replication CMH P values. SNPs with combined P values $< 5\times10^{-8}$ were considered to have genome-wide significant evidence for association27. SNPs with genome-wide significant evidence for association in the UC GWAS and CMH P values < 0.05 in the replication samples were considered to have confirmed association with UC.

Parents of UC GWAS and replication sample cases with available DNA samples were also genotyped, and affected offspring trios were analyzed using the transmission/disequilibrium test15 (see Supplementary Table 1 online).

Conditional analysis

We were interested in whether one or more causal SNPs might explain the observed association in regions showing genome-wide significant association with UC. To explore this question we fit a conditional logistic model, with pairs of SNPs in the model, conditional on ancestry and gender. P values for the SNPs are interpretable as the residual variation explained by the SNP, conditional on the inclusion of the other SNP. Significant residual association signal was defined as P < 0.05 in the conditional analysis. We also computed pairwise r^2 values in our GWAS control data for the same pairs of SNPs.

Demarcation of regions containing genome-wide significant association signals

We demarcated regions containing genome-wide significant association signals by first defining the set of SNPs with $r^2 > 0.5$ to all independent SNPs showing genome-wide significant association using a 1 Mb window of HapMap CEU data23 centered on the index SNP. SNPs were considered to be independent if they had a $r^2 < 0.3$ in our GWAS control data. Recombination hotspots28 flanking these sets of SNPs were set as the boundaries of regions containing genome-wide significant association signals.

Power analysis

A post hoc power analysis was computed for a two-stage study design with parameters similar to our study design using the CaTS software26 (see Supplementary Table 5 online).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

The National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) IBD Genetics Consortium is funded by the following grants: DK062431 (S.R.B.), DK062422 (J.H.C.), DK062420 (R.H.D.), DK062432 (J.D.R.), DK062423 (M.S.S.), DK062413 (K.D.T.), and DK062429 (J.H.C.). The authors would also like to acknowledge additional support from the Atran Foundation (S.R.B.), Board of Governor's Chair in Medical Genetics at Cedars-Sinai Medical Center (J.I.R.), Bohmfalk Funds for Medical Research (J.H.C.), Burroughs Wellcome Medical Foundation (J.H.C.), Crohn's and Colitis Foundation of America (C.A., T.M.B., S.R.B., J.H.C., R.H.D., J.D.R.), Crohn's and Colitis Foundation of Canada (M.S.S.), Feintech Chair in Immunobiology (S.R.T.), Gale and Graham

Wright Research Chair in Digestive Diseases (M.S.S.), Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center (T.M.B., S.R.B.), National Institutes of Health grants RR00052 (S.R.B.), DK077905 (C.A.), DK068112 (J-P.A.), DK072373 (J.H.C.), RR024139 (J.H.C.), DK076025 (R.H.D.), DK064869 (J.D.R.), MH057881 (K.R.), DK046763 (J.I.R., S.R.T., K.D.T.), and RR00425 (K.D.T.), the Rainin IBD Genetics Research Fund (J-P.A.), and W. Buford Lewis family (S.R.B.).

Thanks to Lisa Wu Data, Joann Fultz, and Joanne Stempak for coordinating study subject recruitment, to Angelo Andriulli, M.D. and Orazio Palmieri, Ph.D. for providing clinical data, and to Jing Lian, Anthony Liew and Houman Khalili for technical support.

REFERENCES

1. Xavier RJ, Podolsky DK. Unravelling the pathogenesis of inflammatory bowel disease. Nature. 2007; 448:427–434. [PubMed: 17653185]
2. Duerr RH, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science. 2006; 314:1461–1463. [PubMed: 17068223]
3. Libioulle C, et al. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. PLoS Genet. 2007; 3:e58. [PubMed: 17447842]
4. Rioux JD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet. 2007; 39:596–604. [PubMed: 17435756]
5. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]
6. Parkes M, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. Nat Genet. 2007; 39:830–832. [PubMed: 17554261]
7. Barrett JC, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet. 2008; 40:955–962. [PubMed: 18587394]
8. Fisher SA, et al. Genetic determinants of ulcerative colitis include the ECM1 locus and five loci implicated in Crohn's disease. Nat Genet. 2008; 40:710–712. [PubMed: 18438406]
9. Franke A, et al. Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. Nat Genet. 2008; 40:713–715. [PubMed: 18438405]
10. Kugathasan S, et al. Loci on 20q13 and 21q22 are associated with pediatric-onset inflammatory bowel disease. Nat Genet. 2008; 40:1211–1215. [PubMed: 18758464]
11. Franke A, et al. Sequence variants in IL10, ARPC2 and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet. 2008
12. Stokkers PC, Reitsma PH, Tytgat GN, van Deventer SJ. HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis. Gut. 1999; 45:395–401. [PubMed: 10446108]
13. Luca D, et al. On the use of general control samples for genome-wide association studies: genetic matching highlights causal variants. Am J Hum Genet. 2008; 82:453–463. [PubMed: 18252225]
14. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]
15. Ewens WJ, Spielman RS. The transmission/disequilibrium test: history, subdivision, and admixture. Am J Hum Genet. 1995; 57:455–464. [PubMed: 7668272]
16. Makarova KS, Aravind L, Koonin EV. A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and Chlamydia pneumoniae. Trends Biochem Sci. 2000; 25:50–52. [PubMed: 10664582]
17. Murakami M, et al. Arachidonate release and eicosanoid generation by group IIE phospholipase A(2). Biochem Biophys Res Commun. 2002; 292:689–696. [PubMed: 11922621]
18. Suzuki N, et al. Structures, enzymatic properties, and expression of novel human and mouse secretory phospholipase A(2)s. J Biol Chem. 2000; 275:5785–5793. [PubMed: 10681567]
19. Joazeiro CA, Weissman AM. RING finger proteins: mediators of ubiquitin ligase activity. Cell. 2000; 102:549–552. [PubMed: 11007473]
20. Schoenborn JR, Wilson CB. Regulation of interferon-gamma during innate and adaptive immune responses. Adv Immunol. 2007; 96:41–101. [PubMed: 17981204]

21. Aujla SJ, Dubin PJ, Kolls JK. Th17 cells and mucosal host defense. Semin Immunol. 2007; 19:377–382. [PubMed: 18054248]
22. Sugimoto K, et al. IL-22 ameliorates intestinal inflammation in a mouse model of ulcerative colitis. J Clin Invest. 2008; 118:534–544. [PubMed: 18172556]
23. Frazer KA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007; 449:851–861. [PubMed: 17943122]
24. Dixon AL, et al. A genome-wide association study of global gene expression. Nat Genet. 2007; 39:1202–1207. [PubMed: 17873877]
25. Teller IC, Beaulieu JF. Interactions between laminin and epithelial cells in intestinal health and disease. Expert Rev Mol Med. 2001; 3:1–18. [PubMed: 14585148]
26. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. Nat Genet. 2006; 38:209–213. [PubMed: 16415888]
27. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008; 32:381–385. [PubMed: 18348202]
28. Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. A fine-scale map of recombination rates and hotspots across the human genome. Science. 2005; 310:321–324. [PubMed: 16224025]

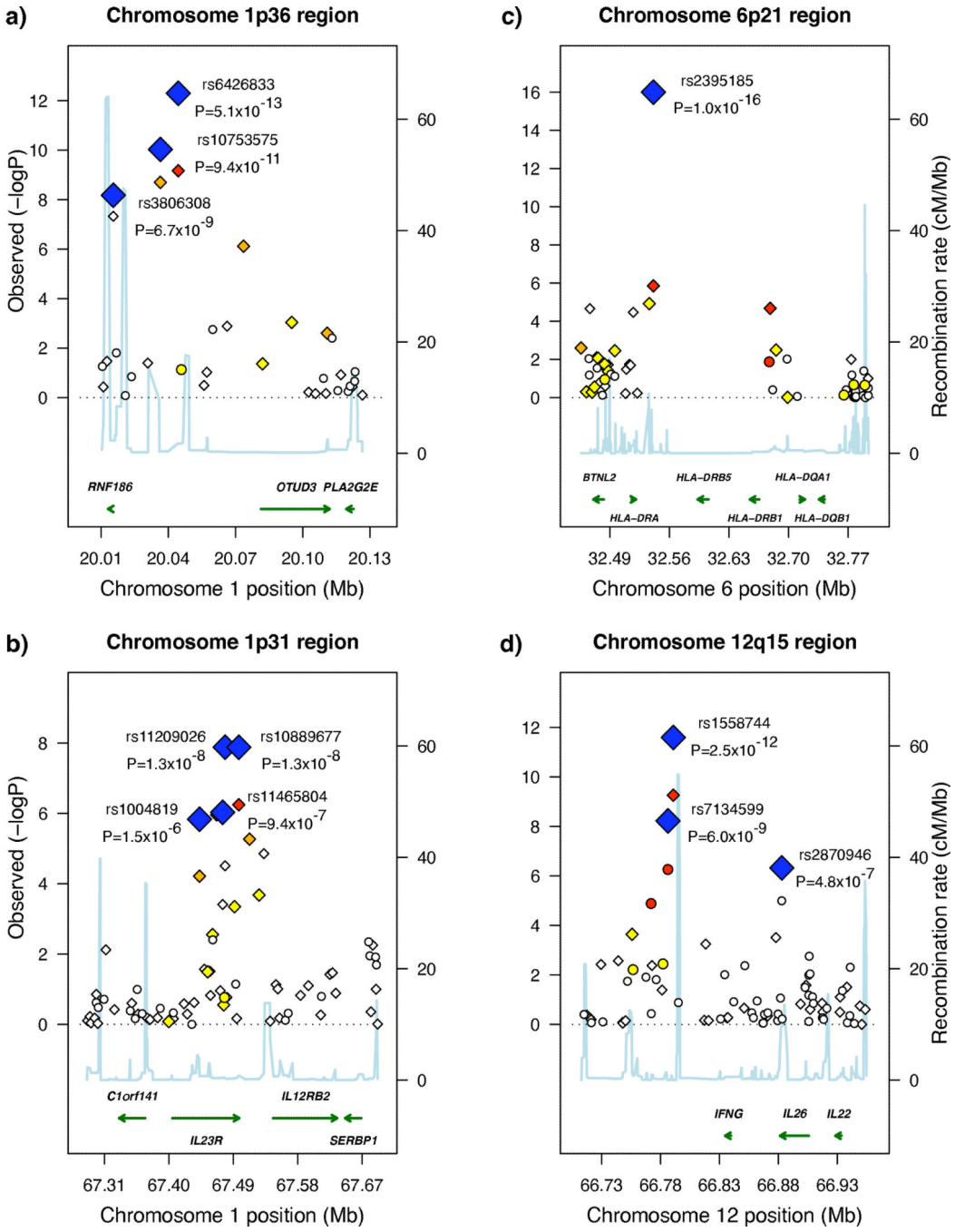

Figure 1.

Association signals and recombination rates for loci demonstrating genome-wide significant association with UC on chromosomes a) 1p36, b) 1p31, c) 6p21 and d) 12q15. The vertical axis scale on the left side of each plot represents the –log10 of the P values. The vertical axis scale on the right side of each plot represents the recombination rate in cM/Mb. The UC GWAS –log10 GEM P values are plotted as small diamonds (all arrays shared SNPs GEM dataset, see Supplementary Table 1 online) and small circles (HumanHap550 only SNPs GEM dataset, see Supplementary Table 1 online). Linkage disequilibrium (r^2) in the GWAS

control data to the single most significantly associated regional SNP is color-coded (red: r^2 greater than 0.8; orange: r^2 0.5−0.8; yellow: r^2 0.2−0.5; white: r^2 less than 0.2). Large blue diamonds with corresponding rs number and P value labels represent combined UC GWAS and replication case-control significance estimates. Light blue lines represent the recombination rates. Conditional analyses (see Supplementary Table 3 online) of replicated markers (large blue diamonds) suggest the presence of independent association signals on chromosome 1p36 and on chromosome 12q15 separated by recombination hotspots.

Table 1

ned association evidence for four regions showing combined genome-wide significant association with UC

		UC GWAS Sample				Replication Samples						Combined	
oci	Minor Allele	MAF Cases	MAF Controls	UC GWAS GEM P	UC GWAS GEM OR	North American Cases MAF	North American Controls MAF	Italian Cases MAF	Italian Controls MAF	Replication CMH P	Replication CMH OR	Combined P	Combined OR
	A	0.30	0.37	4.7×10^{-8}	0.70	0.33	0.38	0.26	0.29	6.2×10^{-3}	0.85	6.7×10^{-9}	0.78
	G	0.31	0.39	2.0×10^{-9}	0.68	0.34	0.39	0.31	0.34	1.7×10^{-3}	0.83	9.4×10^{-11}	0.76
	G	0.37	0.46	6.8×10^{-10}	0.68	0.39	0.46	0.33	0.38	2.3×10^{-5}	0.77	5.1×10^{-13}	0.73
	A	0.35	0.30	6.1×10^{-5}	1.30	0.34	0.30	0.39	0.35	1.4×10^{-3}	1.22	1.5×10^{-6}	1.25
	C	0.039	0.071	3.9×10^{-4}	0.60	0.038	0.066	0.045	0.062	1.4×10^{-4}	0.61	9.4×10^{-7}	0.61
	A	0.037	0.070	3.1×10^{-5}	0.55	0.036	0.067	0.040	0.060	1.9×10^{-5}	0.57	1.3×10^{-8}	0.56
	A	0.36	0.30	5.7×10^{-7}	1.38	0.36	0.31	0.38	0.35	1.0×10^{-3}	1.22	1.3×10^{-8}	1.29
	A	0.24	0.33	1.4×10^{-6}	0.72	0.21	0.32	0.17	0.23	3.6×10^{-13}	0.61	1.0×10^{-16}	0.66
	A	0.44	0.36	5.4×10^{-7}	1.55	0.42	0.38	0.41	0.35	4.8×10^{-4}	1.23	6.0×10^{-9}	1.32
	A	0.48	0.38	5.5×10^{-10}	1.47	0.45	0.41	0.45	0.37	1.5×10^{-4}	1.26	2.5×10^{-12}	1.35
	G	0.10	0.055	1.0×10^{-5}	2.01	0.079	0.062	0.12	0.085	2.6×10^{-3}	1.37	4.8×10^{-7}	1.54